Anomaly detection for sparse data

A framework based on PU-Learning and GAN's

Andrew Shields†
Department of Computing
Institute of Technology Tralee
Co Kerry Ireland
andrew.shields@staff.ittralee.ie

Ted Scully
Department of Computing
Cork Institute of Technology
Cork Ireland
ted.scully@cit.ie

ABSTRACT

Deep Learning is a machine learning method based on neural network architectures with multiple layers of processing units. It has been successfully applied to a number of data mining problems, particularly in the areas of image recognition and natural language processing. Deep learning for anomaly detection is an active and ongoing research topic. Supervised anomaly detection methods emphasise learning feature representations from known anomalies types. However, supervised approaches do not scale well to large data sets and may be unable to generalise well to unknown anomaly types. To counteract this, unsupervised learning techniques are often implemented which operate without labelled anomaly data, such unsupervised approaches do not utilise the sparse amount of normal data which is often easily available. This paper introduces a novel anomaly detection framework to address these problems. Instead of learning anomaly representations from known anomaly types or from unsupervised models, our method leverages a small number of labelled normal (positive) data instances as well as unlabelled instances for training using Positive and Unlabelled (PU) learning as a pre training step to an anomaly detector based on Generative Adversarial Networks (GAN) model. Initial comparative analysis was undertaken between the proposed approach and a one class SVM.

CCS CONCEPTS

• Computing methodologies → Anomaly detection; Deep learning; Semi-supervised learning settings.

KEYWORDS

Anomaly Detection, Generative Adversarial Networks (GAN), PU Learning.

DLP-KDD'20, August 24, 2020, San Diego, CA, USA
 978-1-4503-0000-0/18/06...$15.00
https://doi.org/10.1145/1234567890

1 INTRODUCTION

Anomalies may be defined as data instances which deviate significantly from the normal expected behaviours for the dataset, anomaly detection may therefore be defined as the recognition of these deviations. Anomaly detection has important applications across a number of domains including detecting network attacks in cybersecurity, fraudulent transactions in finance, and diseases in healthcare.

Traditionally statistical methods such as control charts have been applied to anomaly detection. However, these methods are unable to deal with Big Data generated by modern systems. And research has begun to exploit machine learning to overcome these shortcomings. Multiple anomaly detection methods have been proposed, but there are many challenges. [1], [2] These can be complicated by the difficulty in acquiring labelled anomaly data for training, which is often expensive or impossible to obtain.

Anomalies often demonstrate different interclass behaviours, meaning that anomalies types may significantly differ from each other, this poses significant challenges to supervised classification which assumes that data instances within each class are similar to each other. Furthermore, anomalies may sometimes display dynamic behaviour, where the normal behaviour evolves. Therefore, there are often Inaccurate or "soft" boundaries between the anomalous and normal behaviour. This implies that an anomaly detection system should ideally incorporate a degree of continual learning in their frameworks to overcome these challenges.

Deep anomaly Detection (DAD) methods have been proposed as a solution and have been shown to be generally more accurate than semi-supervised and unsupervised models. However, they still require accurate lager amounts of normal and anomalous instances for training. As a result, hybrid semi supervised one-class learning methods become more appealing.

Inspired by these challenges, we propose a novel advancement to the current two step anomaly detection approach. Firstly, incorporating a PU pre-training stage for input, which is a semi-supervised learning method requiring no labelling of anomalies in the training process. Secondly, we incorporate GAN model into the framework to tackle the task of anomaly detection, the generator and discriminator are used to capture the temporal correlation of time series distributions. This differs from current hybrid models in that the modified PU algorithm is used for pre-training while the deep learning GAN is used for the feature extraction and anomaly

detection, current hybrid systems use the deep learner as the pre-trainer and the standard technique as the anomaly detector. The advantage of our approach is that it will operate with only small amounts of normal labelled data while utilising the readily available unlabelled data.

We evaluate the potential of this framework by performing a series of preliminary experiments. To evaluate the performance for the unbalanced problem, the normal class in each dataset has a fixed size, but the size of the anomaly class is decreased, thereby increasing the level of imbalance in the dataset.

2 RELATED WORKS

2.1 Deep Anomaly Detection

Anomaly detection [3] is a technique used to identify unusual patterns whose behaviour does not conform to expected normal behaviour. Much research has been undertaken on the anomaly detection problem using supervised and unsupervised methods. [4], [5], [6], [7], [8].

Due to their nature, labelled anomaly samples are rare occurrences and are usually very difficult to obtain. Therefore, unsupervised learning techniques are a common approach used to solve anomaly detection problems these generally fall into two camps unsupervised methods [9] and one-class classification methods [10].

Deep anomaly detection can capture more complex feature interactions than traditional shallow methods [11] the two most common generative approaches are Variational Autoencoders (VAE) [12], [13] and Generative Adversarial Networks (GAN) [11]. A variant of GAN architecture known as Adversarial autoencoders (AAE) [14] uses adversarial training to impose an arbitrary prior on the latent code learned within hidden layers of autoencoder to learn the input distribution effectively. Leveraging this ability of learning input distributions, several Generative Adversarial Networks-based Anomaly Detection (GAN-AD) frameworks are shown to be effective in identifying anomalies on high dimensional and complex datasets , [15], [16], [11].

Generative Adversarial Networks (GANs) trained in semi-supervised learning mode have shown great promise, especially when there are very few labelled data instances for training. However, GAN's are prone to over-fitting and more traditional methods have been shown to perform better where there is only a smaller number of anomalies. [17]. Therefore, deep hybrid models which employ two step learning are interesting where the representative features learned within deep models are input to more traditional algorithms and are shown to produce state-of-the-art results [18], [19].

2.2 Positive and Unlabelled (PU) learning for anomaly detection

Positive and Unlabelled (PU) Learning is a semi-supervised learning method. PU learning refers to the problem where only labelled positive data and unlabelled data are provided in the training process.

The labelled positive set is denoted as P and unlabelled set is denoted as U. The class prior π is defined as the probability of positive instance occurring. Some work on theoretical conclusions for PU learning [20] compare PU learning against training when all labels are provided. They prove that given infinite unlabelled data, PU learning problems can be solved by cost sensitive learning [21].

Standard PU learners typically assume that a known prior distribution is available and generally operate on balanced distribution between positive and negative. However, for anomaly detection, the positive (normal) data is generally the majority class. Thus, conventional PU learning methods cannot be directly applied to an anomaly detection problem. The literature discusses the proportion of positive instances in unlabelled set, du Plessis et al. [21] propose the use of a linear model for positive and unlabelled problem. However, the linear model is still not sufficient for anomaly detection as inter-class similarities are common. To address this issue a self-learning process may be incorporated by iteratively extracting reliable positive and negative instances from the unlabelled set A boosting-like procedure may be exploited during self-learning to improve the performance of individual classifiers.

3 METHOD

3.1 Overview

This section gives an overview of our developed framework it starts by discussing the operation of the PU learner as a pretraining stage and how bagging may be used to improve the performance of a PU learner in discriminating positive from negative data points. Secondly, we discuss how the GAN model is incorporated into the framework in the training stage. The purpose here is to utilise the discriminator and generator to converge towards the distribution of the positive samples included in the unlabelled dataset. Finally, we discuss the development of a GAN anomaly detector utilising the model developed in the training stage. Each stage of the framework is described in detail below and in figure 1.

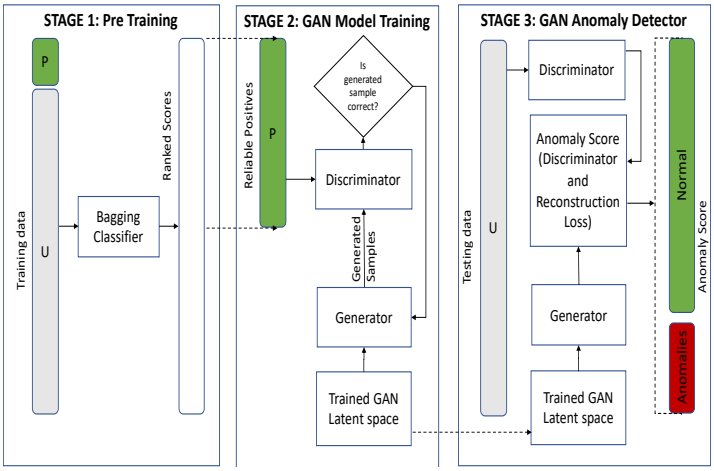

Figure 1 PU GAN anomaly detection framework.

3.2 Stage 1: Pre-Training with PU bagging

The starting point for our framework is the PU learning classifier, which iteratively learns to discriminate positive from unlabelled samples. Many binary classification methods may be used, bootstrap aggregating or "bagging" is a commonly used technique [22]. The final prediction is calculated by aggregating all the individual predictions from each of the base learners [23].

A source of instability in PU learning is the percentage of positive examples in our unlabelled dataset. This is especially problematic for anomaly detection tasks as we expect U to consist primarily of positive samples, negative samples being rare as they are anomalies. Aggregating the classifiers by bagging will induce a large variability in the classifiers. Mordelet and Vert describe this approach [24]. According to the authors, "the method can match and even outperform the performance of state-of-the-art methods for PU learning. For our framework the following steps are performed by PU bagging during Pre-training

1. Create a training set by combining all positive data points with a random sample from the unlabelled points, with replacement.
2. Build a classifier from this "bootstrap" sample, treating positive and unlabelled data points as positives and negatives, respectively.
3. Apply the classifier to the unlabelled data points that were not included in the random sample (from step 1) – hereafter called OOB ("out of bag") points – and record their scores.
4. Repeat the three steps above multiple times and finally assign to each point the average of the OOB scores it has received.

3.3 Stage 2: Model Training with GAN

A GAN generator and discriminator are created. The generator generates fake data sequences as its inputs, and passes the generated examples to the discriminator, which will try to distinguish the generated data sequences from the actual "real" training data sequences. The parameters of the discriminator and generator are updated based on the outputs of the discriminator. This enables the discriminator to assign labels to both real and fake data. After a number of iterations, the generator will have captured the distributions of the training data.

To enable the GAN to capture the relevant dynamics of the data, an LSTM network with depth 3 and 100 hidden units is used for the generator. The network for the discriminator is relatively simpler with 100 hidden units and depth 1. The choice of these settings is directed by the discussion in [25]

3.4 Stage 3: Anomaly Detector with GAN

In this final stage; the previously trained discriminator and generator will be used for the anomaly detection task. The discriminator is used to classify the testing samples which are mapped back into the latent space. Reconstruction loss is calculated established on the variance between the reconstructed testing samples by the generator and the genuine testing samples. At the same time, the testing samples are also fed to the trained discriminator to compute the discrimination loss.

GAN's have an advantage in that the discriminator and the generator are trained concurrently to represent the normal variability for identifying anomalies. Based on [15], the GAN-based anomaly detection consists of the following two parts:

1. Discrimination-based Anomaly Detection; the trained discriminator can distinguish anomalies from real data with high sensitivity, it serves as a direct tool for anomaly detection.
2. Reconstruction-based Anomaly Detection; the trained generator is capable of generating realistic samples, is actually a mapping from the latent space to real data space and can be viewed as a model that reflects the normal data's distribution.

Using the generator and discriminator the two losses are combined to assign an anomaly score to each record this allows us to detect potential anomalies in the data.

4 EXPERIMENTAL SETUP

We have performed Initial comparative analysis of our proposed PU-GAN framework. Through various experiments we show the performance trends of our anomaly detection framework against an OCSVM classifier over various datasets.

4.1 Data Preparation

The purpose of using artificial data is to create an idealised data distribution on which we can concretely test our frameworks performance The synthetic dataset is generated using the SciKit-Learn library in python, the make_moons methods was used to generate two interleaving half circles, an equal number of anomalies and positives data points are present.

For the KDDCUP99 dataset two versions of the dataset are used. Firstly, a highly labelled version of the dataset is used containing, 562387 samples predominantly consisting of positive samples. The second version of the KDDCUP99 dataset is purposely manipulated so that a large number training dataset is unlabelled. PCA was used to reduce the number of attributes from 35 to 6, this is a similar methodology proposed by [26].

We use the metrics, namely Precision, Recall, and F1 scores, to evaluate the anomaly detection performance.

Table 1 Dataset composition

Item	Moons	KDDCUP99 (high labelling)	KDDCUP99 (low labelling)
Attributes	2	6	6
Known Positives	3492	556700	8434
Training size	70000	562387	562387
Testing size	30000	494021	494021
Anomalies in testing	15000	911	911

5 RESULTS

This section discusses our preliminary results from our framework. For comparison on the anomaly detection performance, we also applied a One Class Support Vector Machine (OCSVM) which is a popular unsupervised anomaly detection method on the datasets. It is worth noting that the OCSVM was unable to model

the larger KDDCUP99 dataset, this would suggest the OCSVM may be unsuitable for larger datasets. Future work will fully evaluate the potential for the system. In Table 2, we show the best performance by the unsupervised OCSVM and our PU-GAN. We focus on the results chosen by the F1 score, this is because this measure balance precision and recall.

Table 2 Anomaly detection results for different datasets.

Dataset	Method	Precision	Recall	F1 Score
KDDCUP99 (high labelling)	PU-GAN	0.821	0.925	0.870
KDDCUP99 (low labelling)	PU-GAN	0.816	0.647	0.722
	OCSVM	0.729	0.803	0.716
Moons	PU-GAN	1.0	0.995	0.997
	OCSVM	0.85	0.847	0.847

PU-GAN performed better than the OCSVM for the moons's dataset. This is a balanced dataset with low dimensionality and clear boundaries between the two classes. We also applied PU-GAN to the two versions of the KDDCUP99 dataset. On the dataset with the high number of labelled positives for training, PU- GAN can reach 0.87 F1 score with precision larger than 82% and recall higher than 92%. One point of note is that we have performed our classification only using partially labelled positive records. Testing was also performed on the KDDCUP99 datasets with many of the records in the train set unlabelled. This introduced imbalance to the training data to reflect a more realistic anomaly detection scenario, in this regard our framework still works well and performs better than the baseline OCSVM.

6 SUMMARY AND DISCUSSION

To date we have developed a hybrid solution by using both normal (positive) and unlabelled data for semi-supervised anomaly detection. Particularly, we introduce a new hybrid framework based on PU learning for pre-training in combination with a GAN to detect anomalies. We extend previous PU learning methods to better address the unbalanced class problem, which is typical for anomaly detection, and handle multiple unknown anomaly types.

Our framework has been shown to learn the anomaly classifier incrementally from only the partially labelled positive data and unlabelled data. Preliminary experimental results show that our method performs well under different class priors and different proportions of given positive classes.

These preliminary experiments have only compared with one-class SVM on small datasets. We intend to conduct more complete tests against more state-of-the-art methods published in the area of semi-supervised or unsupervised anomaly detection with public datasets for benchmarking

More precisely we intend to continue and expand on this research in several ways. Firstly, the application of more robust evaluation metrics will be investigated and applied to the results. Secondly, the framework will be evaluated on datasets with varying levels of class imbalance. Finally, we intend to apply the framework to a challenging anomaly detection use case to highlight the application of such a system in a real-world scenario.

REFERENCES

[1] G. Pang, C. Shen, and A. Van Den Hengel, "Deep anomaly detection with deviation networks," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 353–362, 2019.

[2] L. Ruff *et al.*, "Deep One-Class Classification," 2018.

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. September, pp. 1–58, 2009.

[4] C. Zhou and R. C. Paffenroth, "Anomaly Detection with Robust Deep Autoencoders," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17*, 2017, pp. 665–674.

[5] J. Li, W. Pedrycz, and I. Jamal, "Multivariate time series anomaly detection: A framework of Hidden Markov Models," *Appl. Soft Comput.*, vol. 60, pp. 229–240, Nov. 2017.

[6] S. Rajasegarar, C. Leckie, J. C. Bezdek, and M. Palaniswami, "Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 3, pp. 518–533, Sep. 2010.

[7] A. Kulkarni *et al.*, "A Real-Time Anomaly Detection Framework for Many-Core Router through Machine Learning Techniques Real-Time Anomaly De- tection Framework for Many-Core Router through Machine Learning Techniques," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 23.

[8] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, "Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes," *arXiv1609.00866 [cs]*, 2016.

[9] M. Amer, M. Goldstein, and S. Abdennadher, "Enhancing One-class Support Vector Machines for Unsupervised Anomaly Detection."

[10] V. Garcia-Font, C. Garrigues, and H. Rifà-Pous, "A Comparative Study of Anomaly Detection Techniques for Smart City Wireless Sensor Networks," *Sensors*, vol. 16, no. 6, p. 868, Jun. 2016.

[11] D. Li, D. Chen, J. Goh, and S.-K. Ng, "Anomaly Detection with Generative Adversarial Networks for Multivariate Time Series."

[12] S. Suh, D. H. Chae, H. G. Kang, and S. Choi, "Echo-state conditional variational autoencoder for anomaly detection," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, vol. 2016–Octob, pp. 1015–1022.

[13] M. Sakurada and T. Yairi, "Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction," in *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis - MLSDA'14*, 2014.

[14] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial Autoencoders," 2015.

[15] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10265 LNCS, pp. 146–147.

[16] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, 2019.

[17] V. Škvára, T. Pevný, and V. Šmídl, "Are generative deep models for novelty detection truly better?"

[18] S. M. Erfani *et al.*, "Robust Domain Generalisation by Enforcing Distribution Invariance."

[19] C. Wu, Y. Guo, and Y. Ma, "Adaptive Anomalies Detection with Deep Network," *Cogn. 2015, Seventh Int. Conf. Adv. Cogn. Technol. Appl.*, no. c, pp. 181–186, 2015.

[20] G. Niu and M. Sugiyama, "Convex Formulation for Learning from Positive and Unlabeled Data Marthinus Christoffel du Plessis."

[21] M. C. Du Plessis, G. Niu, and M. Sugiyama, "Analysis of Learning from Positive and Unlabeled Data."

[22] L. Breiman, "Bagging predictors - Springer," *Mach. Learn.*, 1996.

[23] L. Breiman, "Statistical modeling: The two cultures," *Stat. Sci.*, 2001.

[24] F. Mordelet and J. P. Vert, "A bagging SVM to learn from positive and unlabeled examples," *Pattern Recognit. Lett.*, vol. 37, no. 1, pp. 201–209, 2014.

[25] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs," 2017.

[26] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S. K. Ng, "MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11730 LNCS, pp. 703–716.